APPENDIX G

POSTMEETING COMMENTS

David E. Burmaster

27 April 1998

Memorandum

To: Moderator, Participants, and Attendees --
Workshop on Selecting Input Distributions for Probabilistic Analyses

Via: Kate Schalk, ERG

From: David E. Burmaster

Subject: Thoughts and Comments After the Workshop in NYC

After much more reading and thinking, I remain staunchly opposed to letting the US EPA and its attorneys set a minimum value for any or all goodness-of-fit (GoF) tests such that an analyst may not use a fitted parametric distribution unless it achieves some minimum value for the GoF test.

In honesty, I must agree that GoF tests are useful in some circumstances, but they are not panaceas, they do have perverse properties, and they will slow or stop continued innovation in probabilistic risk assessment. The US EPA must NOT issue guidance, even though it is supposedly not binding, that sets a minimum value for a GoF statistic below which an analyst may not use a fitted parametric distribution in a simulation.

Here are my thoughts:

1. Re Data

For physiological data, many of the key data sets (e.g., height and weight) usually come from NHANES or related studies in which trained professionals use calibrated instruments to measure key variables (i.e., height and weight) in a clinic or a laboratory under standard conditions for a carefully chosen sample (i.e., adjusted for no shows) from a large population. These studies yield "blue-chip" data at a single point in time. These data, I believe, contain small but known measurement errors across the entire range of variability. At the extreme tails of the distributions for variability, the data do contain relatively large amounts of sampling error. Even with a sample of n = 1,000 people, any value above, say, the 95th percentile contains large amounts of sampling uncertainty. In general, the greater the percentile for variability and the smaller the sample size, the greater the (sampling) uncertainty in the extreme percentiles.

For behavioral and/or dietary data, many key data sets (e.g., drinking water ingestion, diet, and/or activity patterns) often come from 3-day studies in which the human subject recalls events during the previous days without the benefit of using calibrated instruments in a clinic or laboratory and not under standard conditions. Even though the researchers may have carefully selected a statistical sample from a large population, no one can know the accuracy or precision of the "measurements" reported by the subjects. These studies yield data of much less than "blue-chip" quality for a 3-day interval. These data, I believe, contain large and unknown measurement errors across the entire range of variability. At the extreme tails of the

distributions for variability, the data also contain large amounts of sampling error. For a sample with n = 1,000, any value above, say, the 95th percentile contains large amounts of sampling uncertainty above and beyond the large amounts of measurement uncertainty. Again, the greater the percentile for variability and the smaller the sample size, the greater the (sampling) uncertainty in the extreme percentiles.

My conclusion from this? With all sample sizes, certainly with n < 1,000, I think the data are highly uncertain at high percentiles. I think it is inappropriate to eliminate a parametric model that captures the broad central range of the data (say, the central 90 percentiles of the data) just because a GoF test has a low result due to sampling error in the tails of the data. (This observation supports the idea that fitted parametric distributions may outperform EDFs at the tails of the data.) As Dale Hattis has written, use the process to inform the choice of parametric models -- not a mindless GoF test.

2. Re Fitted Parametric Distributions

As is well known:

a 6-parameter model will always fit data better than a 5-parameter model,
a 5-parameter model will always fit data better than a 4-parameter model,
a 4-parameter model will always fit data better than a 3-parameter model, and
a 3-parameter model will always fit data better than a 2-parameter model.

Thus, GoF tests always select models with more parameters than models with fewer parameters.

This perverse behavior contradicts Occam's Razor, a bedrock of quantitative science since the 13th century.

The venerable Method of Maximum Likelihood Estimation (MLE) offers an approach -- not the only approach -- to this problem. First, the analyst posits a set of nested models in which, for example, a n-parameter model is a special case of an (n+1)-parameter model -- and the (n+1)-parameter model is a special case of an (n+2)-parameter model. Using standard MLE techniques involving ratios of the likelihood functions for the nested models, the analyst can quantify whether the extra parameter(s) provide a sufficiently better fit to the data than does one of the simpler models to justify the computational complexity of the extra parameter(s).

3. Re Continued Innovation and Positive Incentives
to Collect New Data and Develop New Methods

Over the last 15 years, the US EPA has issued innumerable "guidance" manuals that have had the perverse effect of stopping research and blocking innovation -- all in the name of "consistency."

In my opinion, our profession of risk assessment stands at a cross-road. The US EPA could specify, for example, all sorts of numeric criteria for GoF tests -- but the casualties would be (i) the continued development of new ideas and methods, especially the theory and practice of "second-order" parametric distributions and the theory and practice of "two-dimensional" simulations, and (ii) the use of expert elicitation and expert judgment.

I again urge the Agency print this Notice inside the front cover and inside the rear cover of each Issue Paper / Handbook / Guidance Manual, etc. related to probabilistic analyses -- and on the first Web page housing the electronic version of the Issue Paper / Handbook / Guidance Manual:

> This Issue Paper / Handbook / Guidance Manual contains guidelines and suggestions for use in probabilistic exposure assessments.
>
> Given the breadth and depth of probabilistic methods and statistics, and given the rapid development of new probabilistic methods, the Agency cannot list all the possible techniques that a risk assessor may use for a particular assessment.
>
> The US EPA emphatically encourages the development and application of new methods in exposure assessments and the collection of new data for exposure assessments, and nothing in this Issue Paper / Handbook / Guidance Manual can or should be construed as limiting the development or application of new methods and/or the collection of new data whose power and sophistication may rival, improve, or exceed the guidelines contained in this Issue Paper / Handbook / Guidance Manual.

References

Burmaster & Wilson, 1996

Burmaster, D.E. and A.M. Wilson, 1996, An Introduction to Second-Order Random Variables in Human Health Risk Assessment, Human and Ecological Risk Assessment, Volume 2, Number 4, pp 892 - 919

Burmaster & Thompson, 1997

Burmaster, D.E. and K.M. Thompson, 1997, Fitting Second-Order Parametric Distributions to Data Using Maximum Likelihood Estimation, Human and Ecological Risk Assessment, in press

P. Barry Ryan

Colleagues-

I read with interest the comments forwarded by Dr. David Burmaster regarding the conference from last week.

I would like to add a few similar words regarding the codification of any specific values for any specific goodness-of-fit (GOF) tests.

GOF tests, by their nature, are very restrictive in affording acceptance of a distribution. For example, the Kolmorgorov-Smirnoff test chooses the largest difference between the observed data and the theoretical ranking and tests using that. Unusual occurrences in data, minor contamination of by other distributions, etc., can cause rejection of distributions that otherwise pass the "duck test" (if it walks like a duck,...)even if one point looks a little more like a pigeon. The GOF test will end up rejecting pretty much everything leaving one with no choice but to use an EDF.

Unfortunately, EDFs are not readily amenable to analyses that lend a lot of insight (cf., Wallace, Duan, and Ziegenfus, 1994). If EPA codifies a fixed value, even in the guise of "guidance" pretty soon no pdf will be safe from legal wrangling.

We spent a long time at the workshop fussing over definitions of representativeness, sensitivity, etc., with little focus on the utility of the techniques. EPA may well be in the difficult position of having to
defend everything from a legal perspective. However, the preoccupation with numbers often comes at the expense of insight. The role of probabilistic assessments is the latter. Our goal is to understand exposure and its influence on health, not to focus on a specific value of a GOF test statistic.

Somewhere in this document should be a statement equivalent to the one often seen in automobile commercials. "The material and techniques contained herein should only be used by professionals familiar with the nuances of the problem at hand and the techniques used, their limitations, and strengths." I object to the cookbook approach to this type of assessments.

I will now step down off my soapbox.

P. Barry Ryan
Professor, Exposure Assessment and Environmental Chemistry
Rollins School of Public Health
Emory University
1518 Clifton Road NE
Atlanta, Georgia 30322
(404) 727-3826 (Voice)
(404) 727-8744 (Fax-Work)
bryan@sph.emory.edu